View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Dokumenten-Publikationsserver der Humboldt-Universität zu Berlin

Susanne Dobratz
Humboldt-Universität zu Berlin
Electronic Publishing Group
University Library/Computer- und Media service
dobratz@cms.hu-berlin.de

Dr. Astrid Schoger
Bayerische Staatsbibliothek (Bavarian State Library)
Digital Library
schoger@bsb-muenchen.de

Digital Repository Certification: A Report from Germany

1 Thoughts about Certification

Certification can be interpreted and implemented in different ways. It always depends on the overall goal to be reached through the certificate and the auditing process. In order to support the development of certain standards it is more important to encourage repositories to use Persistent Identifiers than to exclude them by defining criteria that only a minority of repositories could fulfill. Demanding usage of a special Persistent Identifier (e.g. DOI or URN) would be a typical example here. This kind of certification supports the coaching concept and we regard this as *Soft Certification*. However, a certificate which has the primary objective of ensuring the highest level of trust would exclude the majority of repositories, only admitting those that follow very strict rules. We regard this kind as *Hard Certification*.

Within this article we present 2 approaches from Germany:
The *DINI*[i] *certificate for document and publication repositories* [1] that aims to network document and publication repositories by pushing the use of standards and promoting interoperability. And the work that is being carried out by *nestor*, the Network of Expertise in Long-term STOrage of Digital Resources - A Digital Preservation Initiative for Germany[ii]. The *nestor Working Group on Trusted Repository Certification,* established in December 2004, investigates the standards and methodologies that will ultimately lead to trustworthy digital repositories.

Both certificates will ensure quality and raise trustworthiness by guaranteeing that publication servers and repositories are set up and operated according to certain standards and best practices. This enables interoperability for collaborative publication tasks and long-term preservation.

For *DINI* the primary objective of the guidelines and criteria was to improve interoperability and cooperation between German higher education institutions that run digital repositories and to provide an instrument for the repository operators that could be used to raise the visibility, the recognition and the importance of the digital repository within the university. The certificate shows potential users and authors of digital documents that a certain quality level in operating the repository is guaranteed and that this distinguishes it from common web servers of institutions. In addition, *DINI* sees its certificate as an instrument to support the Open Access concept. This certificate can be

viewed as a "soft certificate", where the coaching idea prevails. It works on the basis of self disclosure by the repositories.

nestor with its certificate, by contrast, aims to document the trustworthiness[iii] of digital repositories not only in higher education institutions but also in national and state libraries and archives, in museums and data centers as well as for data producers and service providers. Trustworthiness is important to potential customers, who are the producers of digital information on the one hand, and the readers of the deposited information on the other hand. Also, cooperation partners and an organization's own management have to be able to rely on the trustworthiness of the digital archives. As Jantz and Giarlo [2] state, *"For repositories of scholarly materials, trust can become a significant long-term barrier and considerably increase the complexity of the digital preservation task"*.

The concept of trust applies not only to technological issues, but also to organizational and cultural aspects. The RLG/OCLC Working Group and the RLG/NARA Task Force expressed this in their report in 2002 [3] [7] and in their criteria catalogue in September 2005 [4]. Trustworthy digital repositories as defined by *nestor* can assure producers of all kinds of digital objects that their content is secured and preserved in a manner that ensures their authenticity and data integrity and also takes various right issues into account. It also provides a certain security level to the end user, meaning that the information is accessible over time and that what he receives from the digital repository is trustworthy in terms of the authenticity of the object, the producer, publication time and place. For the institution itself and its cooperation partners, the certificate guarantees the reliability of the digital archive services, which is a prerequisite for the integration in the overall mission of the institution, and for collaboration on the national or international level.

Therefore *nestor* goes further then *DINI* and aims to establish a three-stage process. This starts with giving *orientation* for the planning and implementation phase by providing checklists with recommendations, standards and best practices. In the second step it enables *qualification* by means of self-evaluation and self-representation improving transparency. The last step is a "hard certificate" attesting to a high level of trust, in which a digital repository has to undergo a certification process by external certification experts, normally based on an internationally standardized criteria catalogue.

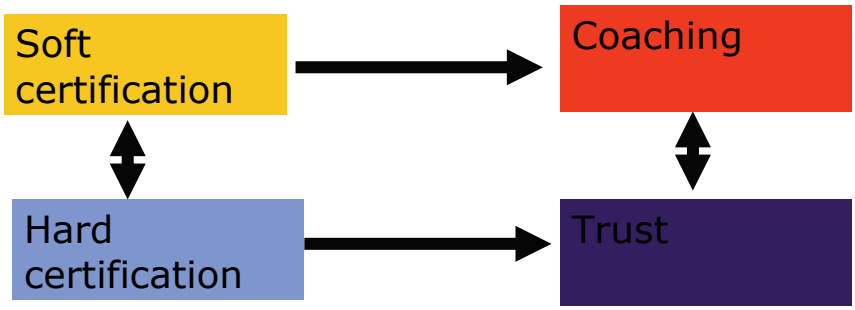

Figure 1: Types of Certification

As the long-term preservation of digital objects is, globally speaking, in its infancy and little experience has been amassed to date, certification is not intended to *„… give a declaration of guarantee for five or fifty years, but to enable institutions to develop strategies in order to cope with the continuous change of information technology in a responsible way*" [5][9]

2 The *DINI* Certificate for Document and Publication Repositories

2.1 About *DINI*: The German Initiative for Networked Information

DINI is the coalition of German infrastructure- or service-institutions: the libraries, the computing centers, the media centers and the scientists, which are represented by the Information and Communication Initiative of their Learned Societies.

DINI is an institution that is committed to initiating and intensifying the regional, nationwide and international collaboration between the infrastructure facilities at universities and to creating recommendations for efficient information services and communication networks in and between universities. A comparable initiative is CNI, the Coalition of Networked Information in the USA, which is a cooperation partner of *DINI*.

DINI operates several working groups, one of which is the Working Group on Electronic Publishing.

2.2 Short Description of the *DINI* Certificate

Installing and institutionalizing document and publication repositories at universities allows them to offer and archive scholarly publications that have been produced at the respective universities to a worldwide audience. This new service by the library and computing center infrastructure institutions helps disseminate the concept of electronic publishing as a new tool for academic work.

"DINI supports this development in order to reach a higher level of scientific and scholarly communication nationwide and internationally and highlights the necessity to network document and publication repositories. Such a network of local publications would complement the dominant (commercial) publications through publishing houses and may therefore also function as a regulating body against the monopolistic tendencies in scholarly publishing. The German Science Council and the Conference of University Rectors in Germany call for the installation of document and publication repositories, and funding institutions such as the Ministry for Science and Technology[iv] and the German Research Foundation[v] support and fund this. It is deemed important to develop this according to international standards and to use proven technology." [1]

Since 1997 we have been developing digital repositories in Germany, mostly funded by the German Research Foundation or the German Ministry for Education and Research. In 2003 *DINI* carried out a survey of German University Repositories and received answers from 47 universities. The kinds of documents stored in these repositories vary from theses and dissertations, monographs, journals, preprints, papers, teaching materials to historical digitalized materials. In this survey *DINI* found out that e.g. only 40 % are working to any kind of policy. The answers to the questionnaire showed a very disappointing use of standards for metadata, interfaces, cataloguing, subject cataloguing or organizational regulations.

Therefore the *DINI* working Group decided to issue a certificate in order to motivate operators of institutional repositories to use appropriate technology and methods. The certificate was launched in 2003 and introduced quality control for servers for the first time. A set of minimum requirements for repositories and their operators which is regarded as mandatory for modern scholarly communication was developed, as well as recommendations highlighting foreseeable developments that might turn into future requirements.

Figure 2: The Logo distributed to certified servers.

2.2.1 Criteria

The *DINI*-criteria are split into two sections. The first section specifies *minimum standards and requirements* that must be met by the document and publication repositories or their operators in order to be awarded the certificate. As ***DINI*** plans continuous adaptation to ensure that the certificate retains its validity in an ever changing environment, the second section lists *recommendations* that, from today's point of view, are likely to be future requirements for the certificate.

The requirements and recommendations cover the following topics:

- Server Policy / Guidelines
- Author support
- Legal issues
- Authenticity and integrity
- Cataloguing
- Access statistics
- Long-term availability

2.2.2 Auditing

A working group[vi] within *DINI* audits the criteria for the *DINI* certificate against international standards and developments and updates them accordingly. For this reason the certificate is issued with a year-of-award stamp. The *DINI* office or an authorized working group is responsible for awarding the *DINI* Certificate – Document and Publication Repository. The certification document acknowledges that the certified repository meets the minimum standards of a *DINI*-certified Document and Publication Repository. A small fee (EUR 50 up to EUR 250) is charged for issuing the *DINI* certificate.

2.3 Experience with the Certification Process

So far, 14 university repositories[vii] have been awarded (or have applied for) the certificate. In discussion, colleges report that the certification procedure has caused local authorities to reflect more deeply about the repository service itself for the first time and to start thinking about their repositories' mission and philosophy. The recommendations and the guidelines were found to be a good orientation in bringing the local repository up to a certain level and in bringing this to the attention of the institution's directorate.

Figure 3: German Repositories with DINI Certificate

3 Certification within *nestor*

3.1 About *nestor*: Network of Expertise in Long-term STOrage of Digital Resources - A Digital Preservation Initiative for Germany

nestor's objective is to create a network of expertise in the long-term storage of digital resources for Germany, comparable to initiatives like the Digital Preservation Coalition in the UK. As the perspective of current and future archive users is vital to the project, the emphasis is put on the long-term *accessibility* of digital resources rather than on pure *preservation* aspects.
nestor wants to strengthen the awareness of this important and urgent topic among the general public and experts by launching a discourse about "Information Life Cycle Management". The project also makes distributed expertise visible and accessible, provides information and encourages communication between all players involved in long-term preservation in Germany. To this end *nestor* has initiated expert reports on different topics (see [9][10],[10],[11]), has set up a web portal on digital long-term preservation for Germany including a subject gateway and an experts' database. *nestor* operates several working groups; one of these is the WG on Trusted Repositories Certification. Others include those on multimedia archiving and preservation policies.

3.2 The *nestor* Working group on Trusted Repositories Certification

The *nestor working group*[viii] wants to encourage the implementation of reliable digital object repositories and to give orientation for repository managers or vendors, to standardize and to enable cooperation between repositories by providing a criteria catalogue and suggesting metrics for the evaluation, taking into account the special German conditions in preparing for the certification process.

The working group consists of selected representatives of various interest groups with a stake in digital long-term archiving: producers and users of digital information, operators of digital long-term archives, memory organizations, and technical experts. To ensure that the work attracts the cooperation of as wide an audience as possible, the group organizes workshops and round table discussions.

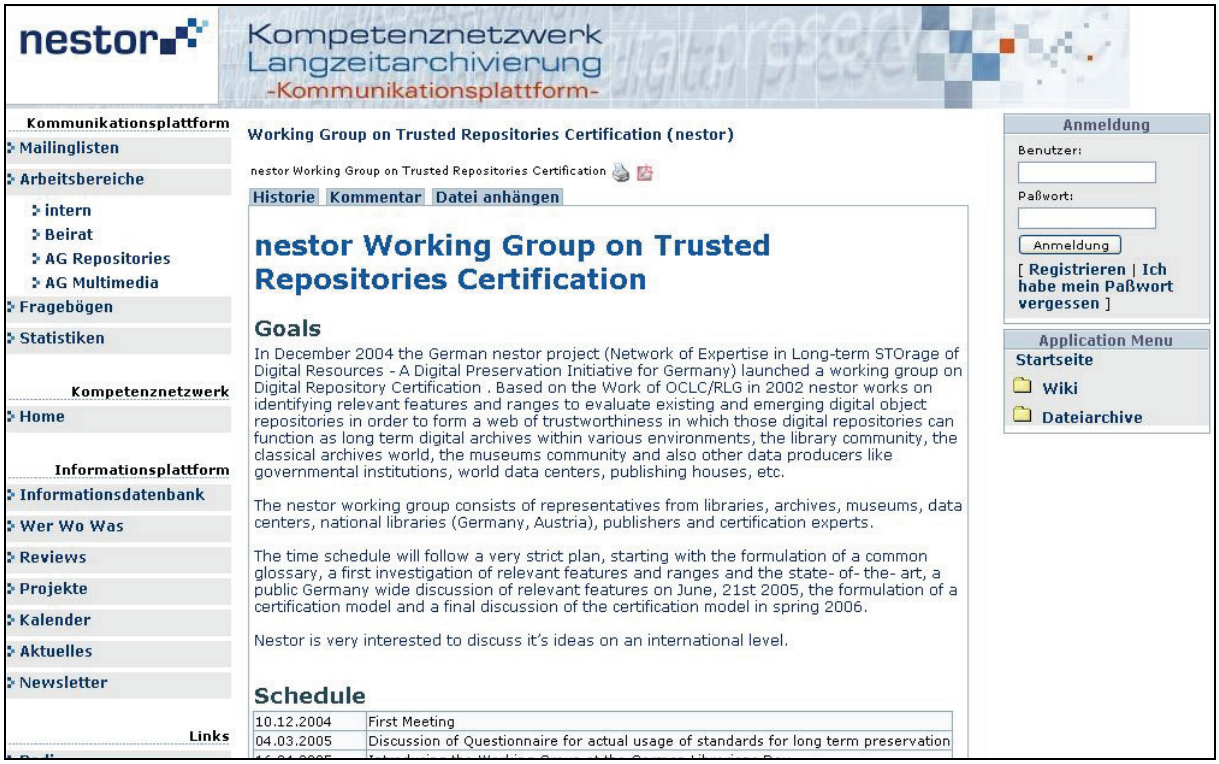

Figure 4: Homepage of the nestor WG on Trusted Repository Certification

The intention is not to reinvent the wheel. It is likely that a catalogue containing the same criteria will result from the joint discussions with the RLG/ NARA Group. It can be organized under different views and undergo specified auditing processes. Important components for the evaluation include local requirements and the different political, legal and financial situations in Germany, Europe and the USA for archives, museums, libraries, data centers etc., and accordingly there is undoubtedly a need for different evaluation schemata taking these special conditions into account.

We are currently discussing several views and our preference at the moment is for the "users" view of such a criteria catalogue. The prime argument for this decision is that the criteria catalogue should function as a guideline for planning, organizing and implementing the requested functionalities. As the realization of the functionalities lies in different hands, it is necessary to give the repository manager a different view and checklist to the systems engineer or the security officer or even the financial manager. Furthermore, the catalogue may be used for self-evaluation. As the transparency increases, producers and end users in particular become interested. Thus we defined four groups:

1. Organizational Issues:
    1.a Internal Organization

1.b Organization of Cooperation with Producers and End Users
2. Technological Issues:
2.a Technical System Management
2.b Technical Object Management

Another view could be one based on the RLG/OCLC attributes.

Criteria Group	Criteria
Internal Organization	Integration of long-term preservation within the overall mission of the institution Preservation policy Documentation of goals, responsibilities, processes, resources Continuity of the digital archive Overall quality management Resource and financial planning
Organization of Cooperation with Producers and End Users	Designated community Preservation policy Selection criteria Collecting guidelines Archiving agreements Interactions with producers Access and usage policy Services offered
Technical System Management	System-related quality management Compliance with technical standards Documentation of technical processes and systems Authenticity and integrity on system level Hard and software environment Feasibility to migrate the contents and the system Flexibility of the system
Technical Object Management	Object-related quality management Metadata for description, preservation, access Controlled vocabulary for metadata Persistent Identifiers Authenticity and integrity on object level Archival formats Long-term availability of the objects

Table 1: nestor Criteria Catalogue

We show here only examples of the main requirements, not the whole criteria catalogue (current working version).

3.3 Survey and Workshop

In order to achieve an evaluation which conforms to current technological and organizational standards, the group performed a survey with representative partners from different sectors: libraries, museums, archives, research institutions, data centers, publishing houses, enterprises, broadcasting stations and a meteorological service. The survey investigated common practices for storing and preserving digital objects. Some results were rather disappointing and showed that a lot of informative work and persuasion has to be done to make the demand for reliable digital objects repositories visible.

As a result, at a workshop in June 2005 the *nestor* WG on Trusted Repositories discussed in four individual groups the important criteria of trusted long-term archives with a wide range of specialist "users", in order to test their applicability and to ensure the relevance of the assessments.

The final discussion revealed that reliability could be achieved in various ways. Firstly, during planning and implementation, a long-term digital archive can follow recommendations, standards and best practices. In a further step, it can become qualified by means of self-representation and improve its transparency to users, producers, cooperation partners and its own management. Furthermore, a long-term digital archive can gain a certificate and, by means of a formal procedure based on strict criteria, establish a high level of trust.

3.4 *nestor*'s evaluation schema for the criteria / metrics

For the survey and the workshop we supplemented the criteria catalogue with examples of typical values, known standards and best practice examples in order to make the questions as precise as possible and to make answering easier.
At the workshop we discussed four main criteria:

- Are there more typical values, best practice examples, standards?
- Is there a consistent set of minimum requirements, or continuative recommendations for these questions?
- How can the degree of performance be measured? Is it a yes / no question or is a differentiated rating scale needed?
- How is this requirement weighted within the overall assessment?

From the results of the survey and the discussion on the workshop we will elaborate a catalogue of criteria, enhanced with metrics for the evaluation of digital repositories, which we hope will be an instrument to provide orientation and for qualifying by self-evaluation, and will formally and stringently lay the groundwork for certification.
There will be a further round table discussion in spring 2006 when we will present and discuss the evaluation schema in depth.

4 Conclusion

We are applying our experience with the *DINI* certification of document and publication repositories and the work done in the *nestor* working group to the international discussion with the RLG/NARA Repository Certification Task Force[ix], the Digital Curation Centre[x], DELOS Network of Excellence on Digital Libraries[xi] and other international activities in order to contribute to the ongoing work on standardization of the criteria catalogue and the certification process.

5 Bibliography

[1] Deutsche Initiative für Netzwerkinformation (*DINI*) Working Group Electronic Publishing: *DINI*-Certificate Document and Publication Repositories, November 2003, *DINI*-Schriften 3-en, urn:nbn:de:kobv:11-10046073

[2] Ronald Jantz and Michael J. Giarlo: Digital Preservation – Architecture and Technology for Trusted Digital Repositories, D-Lib Magazine Vol. 11 No. 6, June 2005, doi:10.1045/june2005-jantz

[3] RLG/OCLC Working Group on Digital Archive Attributes: Trusted Digital Repositories: Attributes and Responsibilities : An RLG-OCLC Report / Mountain View, CA : RLG, 2002 URL: http://www.rlg.org/en/pdfs/repositories.pdf . - [last viewed 28.04.2005]

[4] RLG/NARA Task Force on Digital Repository Certification: Audit Checklist for Certifying Digital Repositories, http://www.rlg.org/en/pdfs/rlgnara-repositorieschecklist.pdf (last viewed 23.09.2005)

[5] Ute Schwens, Hans Liegmann; Die digitale Welt – eine ständige Herausforderung: Rainer Kuhlen, Thomas Seeger und Dietmar Strauch (Eds) In: Grundlagen der

praktischen Information und Dokumentation: Volume 1. *Preprint*, 5th edition. München, Saur, 2004 http://www.langzeitarchivierung.de/downloads/digitalewelt.pdf.
[6] Dobratz, Susanne; Neuroth, Heike: *nestor* : Network of Expertise in Long-term Storage of Digital Resources –A Digital Preservation Initiative for Germany http://www.dlib.org/dlib/april04/dobratz/04dobratz.html, DOI:10.1045/april2004-dobratz
[7] Dobratz, Susanne; Neuroth, Heike; Schoger, Astrid; Strathmann, Stefan: *nestor* – Entwicklungsstand des Kompetenznetzwerks zur Langzeitarchivierung digitaler Ressourcen in Deutschland. Zeitschrift für Bibliothekswesen und Bibliographie, 52 (2005) 3-4 p. 151-162 (see http://www.klostermann.de/)
[8] Henry Gladney: Perspectives on Trustworthy Information, In Digital Document Quarterly (DDQ), Volume 1, Number 2, 1Q2002 , http://home.pacbell.net/hgladney/ddq_1_2.htm
[9] Dirk Witthaut unter Mitarbeit von Andrea Zierer, Arno Dettmers, Stefan Rohde-Enslin: Digitalisierung und Erhalt von Digitalisaten in deutschen Museen: *nestor* Materialien 2: Frankfurt am Main , *nestor* c/o Die Deutsche Bibliothek, 2005 urn:nbn:de:0008-20041223022.
[10] Uwe M. Borghoff u. Mitarb.: Vergleich bestehender Archivierungssysteme: *Nestor* Materialien 3: Frankfurt am Main , *nestor* c/o Die Deutsche Bibliothek und Univ. d. Bundeswehr München Fak. f. Informatik, Inst. f. Softwaretechnologie., 2005 urn:nbn:de:0008-20050117016.
[11] Rechtsanwälte Goebel und Scheller (Bad Homburg v.d.H.): Digitale Langzeitarchivierung und Recht: *nestor* Materialien 1: Frankfurt am Main , *nestor* c/o Die Deutsche Bibliothek, 2004 urn:nbn:de:0008-20040916022.

[i] DINI: Deutsche Initiative für Netzwerkinformation = German Initiative for Networked Information. For more information see http://www.dini.de/.
[ii] For more information on *nestor* see http://www.longtermpreservation.de or [6],[7].
[iii] We use trustworthiness (Vertrauenswürdigkeit) here instead of trusted, because this term is more precise, see Discussion in [8].
[iv] Bundesministerium für Forschung und Bildung
[v] Deutsche Forschungsgemeinschaft
[vi] Members are listed at http://www.*DINI*.de/*DINI*/zertifikat/zertifikat_cons.php
[vii] A complete list of certified servers can be found at http://www.*DINI*.de/*DINI*/zertifikat/zertifiziert.php
[viii] For more information on the *nestor* WG on Trusted Repositories Certification see http://nestor.cms.hu-berlin.de/tiki/tiki-index.php?page=Working+Group+on+Trusted+Repositories+Certification+%28nestor%29
[ix] see: http://www.rlg.org/en/page.php?Page_ID=367
[x] see http://www.dcc.ac.uk/
[xi] see http://www.delos.info/